A 3D DISCRETE CURVELET BASED METHOD FOR SEGMENTING DYNAMIC TEXTURES

Sloven DUBOIS[1,2], *Renaud PÉTERI*[1]

[1]MIA - Laboratoire Mathématiques,
Image et Applications
Avenue Michel Crépeau
17042 La Rochelle, France

Michel MÉNARD[2]

[2]L3i - Laboratoire Informatique
Image et Interaction
Avenue Michel Crépeau
17042 La Rochelle, France

ABSTRACT

This paper presents a new approach for segmenting a video sequence containing dynamic textures. The proposed method is based on a 2D+T curvelet transform and an octree hierarchical representation. The curvelet transform enables to outline spatio-temporal structures of a given scale and orientation. The octree structure based on motion coherence enables a better spatio-temporal segmentation than a direct application of the 2D+T curvelet transform. Our segmentation method is successfully applied on video sequences of dynamic textures. Future prospects are finally exposed.

Index Terms— 2D+T discrete curvelet transform, video segmentation, octree structure, dynamic textures.

1. INTRODUCTION

A moving crowd, rippling water, smoke or grass blown by the wind are different aspects of visual patterns called Dynamic Textures (DT). DT are time-varying patterns with a certain spatial and temporal stationarity. As the real world scenes include a lot of these motion patterns, any advanced video processing application will need to be able to handle DT. Dynamic textures are currently a very active research topic ([1, 2, 3, 4, 5, 6]). Being able to segment these DT spatially and temporally is a highly challenging problem because of their unknown spatial and temporal extend. In this article, we present a method for segmenting DT based on an advanced tool of signal processing, the 3D discrete curvelet transform [7].

The curvelet transform has been designed for improving the limitations of the wavelet transform : while wavelets catches 1D singularities, curvelets can detect structures of other dimensional structures (of co-dimension 1). The discrete curvelet has been recently extended to the third dimension [8, 9].

In this article, time is considered as the third dimension and curvelets are applied to image sequences. These 2D+T curvelets detect surface-like singularities and can then be a promising tool for studying dynamic textures that are often composed of propagating fronts.

This article is organized as follows : section 2 presents the main properties of the 3D discrete curvelet transform. In section 3, the 3D discrete curvelet transform is directly applied to synthetic video sequences and results are discussed. A new method for segmenting a video with dynamic textures is described in section 4. This method is based on a 2D+T curvelet transform and an octree hierarchical representation. The curvelet transform enables to outline spatio-temporal structures of a given scale and orientation. The octree structure based on motion coherence enables a better spatio-temporal segmentation than a direct application of the 2D+T curvelet transform. Our segmentation method is then applied on video sequences containing dynamic textures (section 5). Future prospects are finally exposed.

2. THE 3D DISCRETE CURVELET TRANSFORM

The 3D curvelet transform of a volume $f(\mathbf{x})$, $f \in \mathcal{L}^2(\mathbb{R}^3)$, $\mathbf{x}$ corresponding to 3D coordinates $\mathbf{x} = (x_1, x_2, x_3)$, outputs a collection of coefficients $c(j, \ell, k)$ defined by the following inner product :

$$c(j, \ell, k) := \langle f, \varphi_{j,\ell,k} \rangle = \int_{\mathbb{R}^3} f(\mathbf{x})\overline{\varphi_{j,\ell,k}(\mathbf{x})}d\mathbf{x} \qquad (1)$$

where $\varphi_{j,\ell,k}$ is the curvelet at scale $j \in \mathbb{Z}$, in direction $\ell \in \mathbb{Z}$ and position $k = (k_1, k_2, k_3)$.

Formula (1) can be expressed in the frequency domain as :

$$c(j, \ell, k) := \frac{1}{(2\pi)^2} \int \widehat{f}(\boldsymbol{\omega})\overline{\widehat{\varphi}_{j,\ell,k}(\boldsymbol{\omega})}d\boldsymbol{\omega} \qquad (2)$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ are frequency domain variables.

In this work, the discrete implementation of the curvelet transform is used (see [8]).

$\widehat{\varphi}_{j,\ell,k}$ is defined in the frequency domain by :

$$\widehat{\varphi}_{j,\ell,k}(\boldsymbol{\omega}) = U_{j,\ell}(\boldsymbol{\omega})e^{i\langle x_k^{(j,\ell)}, \boldsymbol{\omega} \rangle} \qquad (3)$$

with $U_{j,\ell}(\boldsymbol{\omega})$ is a discrete frequency window which isolates frequencies near scales j and direction ℓ.

$e^{i\langle x_k^{(j,\ell)}, \boldsymbol{\omega}\rangle}$ represents the translation of the curvelet at position k. This frequency window is expressed by :

$$U_{j,\ell}(\boldsymbol{\omega}) = W_j(\boldsymbol{\omega})V_{j,\ell}(\boldsymbol{\omega}) \qquad (4)$$

where $W_j(\boldsymbol{\omega})$ and $V_{j,\ell}(\boldsymbol{\omega})$ correspond respectively to the radial and angular frequency windows.

The radial window at scale $j > 0$ is expressed by :

$$W_j(\boldsymbol{\omega}) = \sqrt{\Phi_{j+1}^2(\boldsymbol{\omega}) - \Phi_j^2(\boldsymbol{\omega})} \qquad (5)$$

where Φ is defined as the product of one dimensional low-pass windows $\Phi_j(\omega_1, \omega_2, \omega_3) = \phi(2^{-j}\omega_1)\phi(2^{-j}\omega_2)\phi(2^{-j}\omega_3)$. The function ϕ respects $0 \leqslant \phi \leqslant 1$, is equal to 1 on $[-1; 1]$ and 0 on $]-\infty; -2]$ and on $[2; +\infty[$. This window is colored with mean gray on figure 1.

The angular window is defined according to the unit cube face. For example, $V_{j,\ell}(\boldsymbol{\omega})$ is defined relative to the axis ω_1 by :

$$V_{j,\ell}(\omega_1, \omega_2, \omega_3) = \phi\left(2^{j/2}\frac{\omega_2 - \alpha_\ell\omega_1}{\omega_1}\right)\phi\left(2^{j/2}\frac{\omega_3 - \beta_\ell\omega_1}{\omega_1}\right) \qquad (6)$$

For the other cube faces, the definition is similar by exchanging the role of ω_1, ω_2 and ω_3. This window is represented by light gray color on figure 1.

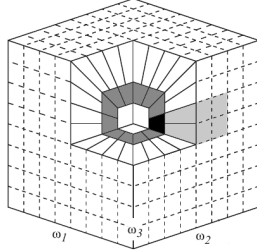**Fig. 1**. Discrete frequency tiling (adapted from [8]). The light and the mean gray colors represent respectively the window $V_{j,\ell}(\boldsymbol{\omega})$ and $W_j(\boldsymbol{\omega})$. The composition of two windows $U_{j,\ell}(\boldsymbol{\omega})$ is colored in black.

The algorithm of the 3D discrete curvelet transform is summarized as follows :
  – Compute the 3D Fourier transform of f.
  – For each scale and each direction, obtain the product $U_{j,\ell}(\boldsymbol{\omega})\widehat{f}(\boldsymbol{\omega})$.
  – Wrap this product around the origin to obtain $\mathcal{W}(U_{j,\ell}\widehat{f})(\boldsymbol{\omega})$. $\mathcal{W}$ is the wrapping operation.
  – Take the inverse 3D Fourier transform to $\mathcal{W}(U_{j,\ell}\widehat{f})$ for collecting the different discrete coefficients $c(j, \ell, k)$.

For more information on the 3D discrete curvelet transform, one can refer to [8, 9].

The 3D discrete curvelet transform was originally designed for three dimensional spatial data (x, y and z). In order to apply this transform to video, we introduce $z = c.t$, where

c is a constant that enables to keep the homogeneity between spatial and temporal variables. The constant c is homogeneous to speed and can be adapted to the considered video.

In the next section, the 2D+T curvelet transform is used for the spatio-temporal segmentation of videos.

3. DIRECT SEGMENTATION USING THE 2D+T DISCRETE CURVELET TRANSFORM

In this section, the 2D+T curvelet transform is applied for extracting DT occurring at different spatio-temporal scales and orientations. This 'direct' method is computed on video of figure 2.(a) that contains a DT orientated with two different directions and of figure 2.(b) where the same DT occurs at two different scales.
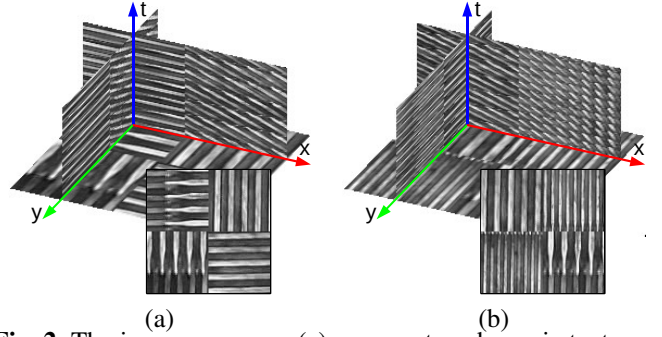

(a)   (b)

Fig. 2. The image sequence (a) represents a dynamic texture (an escalator) with two spatio-temporal directions. One motion is oriented in the xt plane and the other in the yt plane. Video (b) represents the same DT at two different scales (one has a double frequency than the other).

More a frequency with a given orientation ℓ and scale j occurs in a video, higher the energy of the corresponding curvelet will be. In the frequency domain, this phenomenon generates a high response in the window $U_{j,\ell}$. This sector represents the main spatio-temporal direction of the video.

Our method for extracting the main DT in a video is the following :

1. Compute the 2D+T discrete curvelet transform on the whole video.

2. For each direction ℓ at each scale j, compute the energy :

$$E_{j,\ell} = \frac{1}{N_{j,\ell}} \sum_k |c(j, \ell, k)|^2 \qquad (7)$$

where $N_{j,\ell}$ is the number of coefficients located in the wedge (j, ℓ).

3. Compute the inverse 2D+T discrete curvelet transform by keeping the main energy peaks only.

4. Extract a segmentation mask by thresholding in the video domain.

Figure 3 shows the video with the computed mask superimposed. This mask is constructed according to the occurrence of the main direction and scale in the video.
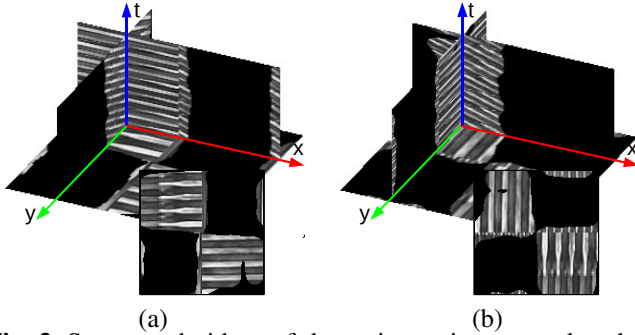

(a)   (b)

Fig. 3. Segmented videos of the main spatio-temporal scale and orientation

Results of figure 3 clearly show that it is possible to segment a video into regions having similar spatio-temporal directions and scales. The segmentation frontiers are however imprecise and only one DT is detected. The next section proposes a method for improving the segmentation.

4. OCTREE DECOMPOSITION USING THE 2D+T DISCRETE CURVELET TRANSFORM

In the previous section, the segmentation was obtained using a 2D+T discrete curvelet transform performed on the whole video. Due to this global computation of the curvelet transform, only the main DT is detected and the extracted spatial and temporal borders are imprecise.

A way to improve this segmentation is to divide the video into several smaller cubes and apply the 2D+T discrete curvelet transform onto each cube. Our method is based on an octree structure for dividing the video and uses the 2D+T discrete curvelet coefficients for the homogeneity criterion. If the video cube is homogeneous in terms of curvelets coefficient energy, the octree division is stopped, otherwise the video cube is again divided into eight subcubes (figure 4). This method tends to subdivide regions located next to the spatio-temporal borders and does not affect regions with homogeneous spatio-temporal frequencies.
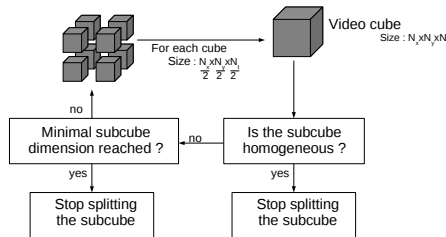

Fig. 4. General principle of the octree structure

This algorithm proceeds as follows :

1: – For a given video cube of size (N_x, N_y, N_t), compute the 3D discrete curvelet transform.
2: – Compute the energy for each orientation ℓ at each scale j.
3: – Energy normalization by a function penalizing spatial directions
4: **if** more then one direction is detected **then**
5:    restart with eight video subcubes of size $\left(\frac{N_x}{2}, \frac{N_y}{2}, \frac{N_t}{2}\right)$.
6: **else**
7:    stop the algorithm here.
8: **end if**

The octree algorithm outputs a tree decomposing the video. To get a spatio-temporal segmentation, the decomposition tree is scanned and similar cubes, in terms of curvelet orientation, are merged. Compared to the method of section 3, several DT can be extracted and a finer spatio-temporal segmentation can be achieved.

5. APPLICATION

The following examples illustrate the whole method on real videos. The first video 5.(a) is a single dynamic texture of rippling water. This dynamic texture is 'disturbed' by a swimming duck. The second video 5.(b) is composed of two dynamic textures having different spatio-temporal characteristics : a waterfall and tree branches waving in the wind.


(a)   (b)

Fig. 5. Original videos. Main spatio-temporal directions are symbolized by arrows.

Results of the spatio-temporal segmentation using the proposed method are displayed on figure 6. The rippling water composed of one spatio-temporal direction (figure 5.(a)) is mainly detected as one homogeneous region (figure 6.(a)). The region containing the duck is considered as an ambiguous area and is thus not colored. One can noticed at the bottom of video 6.(a) an ambiguous area and oversplitted cubes caused by the disturbances created by the wake of the duck.

Video 5.(b) is a difficult case as the two dynamic textures are both transparent and sometimes overlap. One can however see on video 6.(b) that our algorithm was able to distinguish between the two dynamic textures. Some mistakes occur at the border, where ambiguity increases because of some overlaps between the two dynamic textures.
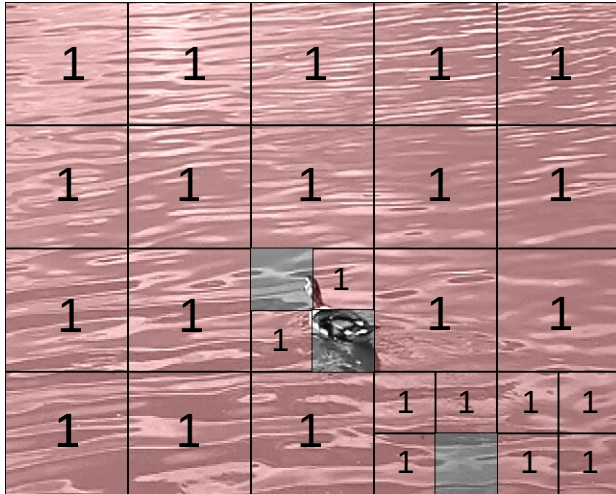
(a)
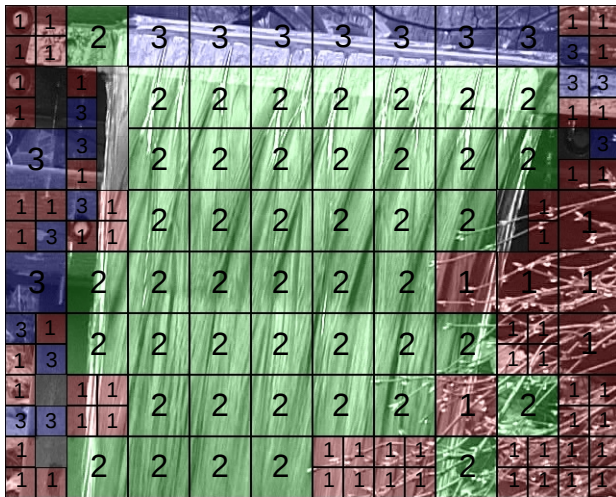
(b)

Fig. 6. Segmentation results of videos shown on figure 5. Each color (red, green and blue, identified respectively by 1, 2 et 3) represents a distinct area. A non colored area corresponds to an ambiguity region. Black segments represent the subcube borders of the octree structure.

Our algorithm can separate dynamic textures in both space and time. Compared to the previous method presented in section 3, several DT can be extracted and the borders are more precisely segmented. One can mention that the precision of the border extraction depends on the shape and size of the smallest octree element.

6. CONCLUSION AND PROSPECTS

This paper explores the use of the 3D curvelet transform for video processing. Two methods for segmenting spatially and temporally image sequences with dynamic textures are exposed. The first method is a direct application of the curvelet transform to the whole video. The second method is based on an octree structure and uses curvelet coefficients as the homogeneity criterion. Applications of this new method on real videos are presented.

Current works aim at modifying the homogeneity criterion for refining the octree subdivisions next to borders. A more geometrically adapted scheme for computing the curvelet transform is also under study.

Segmentation results of this method will be used for the indexation of dynamic textures in large video database [6].

7. REFERENCES

[1] J. Filip, M. Haindl, and D. Chetverikov, "Fast synthesis of dynamic colour textures," in *Proceedings of the 18th IAPR Int. Conf. on Pattern Recognition (ICPR'06)*, Hong Kong, 2006, pp. 25–28.

[2] M. Szummer and R. W. Picard, "Temporal texture modeling," in *Proceedings of IEEE International Conference on Image Processing (ICIP'96)*, 1996, vol. 3, pp. 823–826.

[3] G. Doretto, D. Cremers, P. Favaro, and S. Soatto, "Dynamic texture segmentation," in *Proceedings of Ninth IEEE International Conference on Computer Vision (ICCV'03)*, 2003, vol. 2, pp. 1236–1242.

[4] R. Péteri and D. Chetverikov, "Dynamic texture recognition using normal flow and texture regularity," in *Proceedings of 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'05)*, Estoril, Portugal, 2005, vol. 3523 of *Lecture Notes in Computer Science*, pp. 223–230, Springer.

[5] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence journal (TPAMI'07)*, vol. 6, no. 29, pp. 915–928, 2007.

[6] S. Dubois, R. Péteri, and M. Ménard, "A comparison of wavelet based spatio-temporal decomposition methods for dynamic texture recognition," in *4th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'09)*, Povoa de Varzim, Portugal, 2009, vol. 5524 of *Lecture Notes in Computer Science*, pp. 314–321.

[7] E. Candes and D. Donoho, *Curvelets : A Surprisingly Effective Nonadaptive Representation for Objects with Edges*, pp. 105–120, Curves and Surfaces. Vanderbilt University Press, Nashville, TN, 2000.

[8] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," Tech. Rep., California Institute of Technology, Mar. 2006.

[9] L. Ying, L. Demanet, and E. Candes, "3d discrete curvelet transform," in *Proceedings of the International Society for Optical Engineering (SPIE)*, 2005.